MBMG
Metabarcoding & Metagenomics

Metabarcoding and Metagenomics 2: 1–15
DOI 10.3897/mbmg.2.28963

Research Article

Optimising the detection of marine taxonomic richness using environmental DNA metabarcoding: the effects of filter material, pore size and extraction method

Kristy Deiner[1,2], **Jacqueline Lopez**[2], **Steve Bourne**[3], **Luke E. Holman**[3], **Mathew Seymour**[4], **Erin K. Grey**[5], **Anaïs Lacoursière-Roussel**[6,7], **Yiyuan Li**[2], **Mark A. Renshaw**[2,8], **Michael E. Pfrender**[2], **Marc Rius**[3,9], **Louis Bernatchez**[6], **David M. Lodge**[2,10]

1 *Department of Department of Evolutionary Biology and Environmental Studies, University of Zurich, Zurich, Switzerland.*
2 *Department of Biological Sciences and Environmental Change Initiative, University of Notre Dame, Notre Dame, IN, USA.*
3 *Ocean and Earth Science, National Oceanography Centre Southampton, University of Southampton, UK.*
4 *Molecular Ecology and Fisheries Genetics Laboratory, School of Biological Sciences, Bangor University, UK.*
5 *Division of Science, Mathematics and Technology, Governors State University, IL, USA.*
6 *IBIS (Institut de Biologie Intégrative et des Systèmes), Université Laval, Québec, Canada.*
7 *St. Andrews Biological Station (SABS), Fisheries and Oceans Canada, St.Andrews, NB, Canada.*
8 *Oceanic Institute, Hawaii Pacific University, Waimanalo, HI, USA.*
9 *Centre for Ecological Genomics and Wildlife Conservation, University of Johannesburg, South Africa.*
10 *Cornell University, Atkinson Center for a Sustainable Future and Department of Ecology and Evolutionary Biology, Ithaca, NY, USA.*

Corresponding author: *Kristy Deiner* (alpinedna@gmail.com)

Academic editor: *X. Zhou* | **Received** 8 August 2018 | **Accepted** 19 September 2018 | **Published** 2 November 2018

Abstract

The analysis of environmental DNA (eDNA) using metabarcoding has increased in use as a method for tracking biodiversity of ecosystems. Little is known about eDNA in marine human-modified environments, such as commercial ports, which are key sites to monitor for anthropogenic impacts on coastal ecosystems. To optimise an eDNA metabarcoding protocol in these environments, seawater samples were collected in a commercial port and methodologies for concentrating and purifying eDNA were tested for their effect on eukaryotic DNA yield and subsequent richness of Operational Taxonomic Units (OTUs). Different filter materials [Cellulose Nitrate (CN) and Glass Fibre (GF)], with different pore sizes (0.5 µm, 0.7 µm and 1.2 µm) and three previously published liquid phase extraction methods were tested. The number of eukaryotic OTUs detected differed by a factor of three amongst the method combinations. The combination of CN filters with phenol-chloroform-isoamyl alcohol extractions recovered a higher amount of eukaryotic DNA and OTUs compared to GF filters and the chloroform-isoamyl alcohol extraction method. Pore size was not independent of filter material but did affect the yield of eukaryotic DNA. For the OTUs assigned to a highly successful non-indigenous species, *Styela clava,* the two extraction methods with phenol significantly outperformed the extraction method without phenol; other experimental treatments did not contribute significantly to detection. These results highlight that careful consideration of methods is warranted because choice of filter material and extraction method create false negative detections of marine eukaryotic OTUs and underestimate taxonomic richness from environmental samples.

Key Words

eDNA, 18S ribosomal, seawater, high-throughput-sequencing, metazoan eukaryotes, non-indigenous species

Copyright *Kristy Deiner et al.* This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

PENSOFT

Introduction

Global biodiversity is being redistributed, with negative consequences for humanity (Pecl et al. 2017). Therefore, development of global biodiversity monitoring techniques is needed to more rapidly assess current biodiversity, predict future biodiversity trajectories and improve policy for intervention (Díaz et al. 2015). A technique with broad applications for biodiversity monitoring is the sequencing of environmentally sampled DNA (hereafter eDNA) (Deiner et al. 2017; Stat et al. 2017; Taberlet et al. 2018). However, variations in eDNA methodologies, including sampling, DNA extraction, DNA concentration, library preparation and bioinformatic pipelines, can have a significant effect on the detection of species and downstream ecological inferences (Deiner et al. 2015; Djurhuus et al. 2017; Evans et al. 2017; Leray and Knowlton 2017; Vasselon et al. 2017). No single method has or is likely to emerge as appropriate for all situations due to the ecology of eDNA (*sensu* Barnes and Turner 2016) that is likely to produce spatial and temporal heterogeneity of eDNA in different ecosystems. Indeed, conditions such as pH, temperature and presence of humic substances have been shown to affect the degradation and detectability of eDNA (Seymour et al. 2018; Stoeckle et al. 2017; Tsuji et al. 2017). Given the unlikelihood that one universal method will be ideal for all ecosystems, investigators first need to test the efficacy of a variety of methods to determine what is optimal for their goals and target environment (Goldberg et al. 2016).

Human-modified coastal habitats are key sites for understanding the effects of anthropogenic activities, such as pollution and the introduction of non-indigenous species (NIS) (Floerl and Inglis 2005; Seebens et al. 2013). For example, ballast water tanks and hull fouling of both commercial and recreational ships are prominent vectors dispersing NIS in coastal ecosystems (Gollasch 2002; Lacoursière-Roussel et al. 2012; Molnar et al. 2008; Ruiz et al. 2000; Sylvester et al. 2011). NIS have well-documented negative effects on marine biodiversity and ecosystem services, making bio-surveillance in these environments a high priority (Katsanevakis et al. 2014; Molnar et al. 2008). However, routine monitoring using traditional methods is impractical for many commercial ports around the world. Hence, there is the need for development of an eDNA metabarcoding methodology specific to the study of these coastal environments, which is an attractive alternative given the ease and scalability of sampling compared with traditional methods (Grey et al. 2018; Lacoursière-Roussel et al. 2018).

To date, there is some guidance on suitable filters or DNA extraction techniques for eDNA studies of both fresh and marine environments. Specifically, research has begun to identify optimal methods for filtration and extraction of eDNA in freshwater systems based on qPCR detection of single species and high-throughput sequencing of eukaryotic communities (Deiner et al. 2015; Hinlo et al. 2017; Lacoursière-Roussel et al. 2016; Li et al. 2018; Spens et al. 2017). Only a single study has explored filter material and extraction methods using eDNA metabarcoding to survey marine communities (Djurhuus et al. 2017) , with no existing studies accounting for pore size across different filter material in freshwater or marine systems.

Using seawater sampled from a commercial port, and a factorial experiment (Figure 1), we tested for an effect of filter material, pore size (non-independent of filter material) and extraction method on the DNA yield from eukaryotic environmental DNA (measured after amplification of the V4 region of the18S ribosomal RNA gene) and the correlation that eukaryotic DNA yield has on the detection of Operational Taxonomic Units (OTUs). Our null hypothesis was that filter material, its pore size and the extraction method have no negative effect on the quantity of eukaryotic environmental DNA and number of OTUs detected.

Methods

Experimental design

A power analysis (Zar 1999) was used to determine the required number of experimental replicates per treatment to test the above null hypothesis using the following equation:

$$n = \frac{t^2_{\alpha/2,n-1} CV^2}{(d/\mu)^2}$$

The power analysis calculates sample size (n) or the number of 100 ml experimental replicates, for each response variable; including filter material, pore size and extraction type. We performed the test assuming a 90% confidence that the average DNA concentration after amplification or count of OTUs deviates (d), with a confidence interval (CV) no more than 10% from the true mean (μ) post-amplification DNA concentration or count of OTUs. The true mean is assumed equivalent between replicates. We used the Critical Values calculated from the t distribution assuming a one-tailed test (α=0.05, (α=0.05, Zar 1999) and iteratively solved for sample size until the initial and predicted sample sizes were equivalent. Our statistical power analysis with the above assumptions indicated that we needed at least five replicates per experimental treatment. We therefore strove for ten replicates per experimental treatment (Figure 1) to increase our confidence and compensate for any event during the experiment that would lead to the loss or omission of any replicate.

Sampling site

The Port of Southampton in the United Kingdom is amongst the busiest and oldest commercial ports in Europe (McCutcheon 2008). The port is protected from

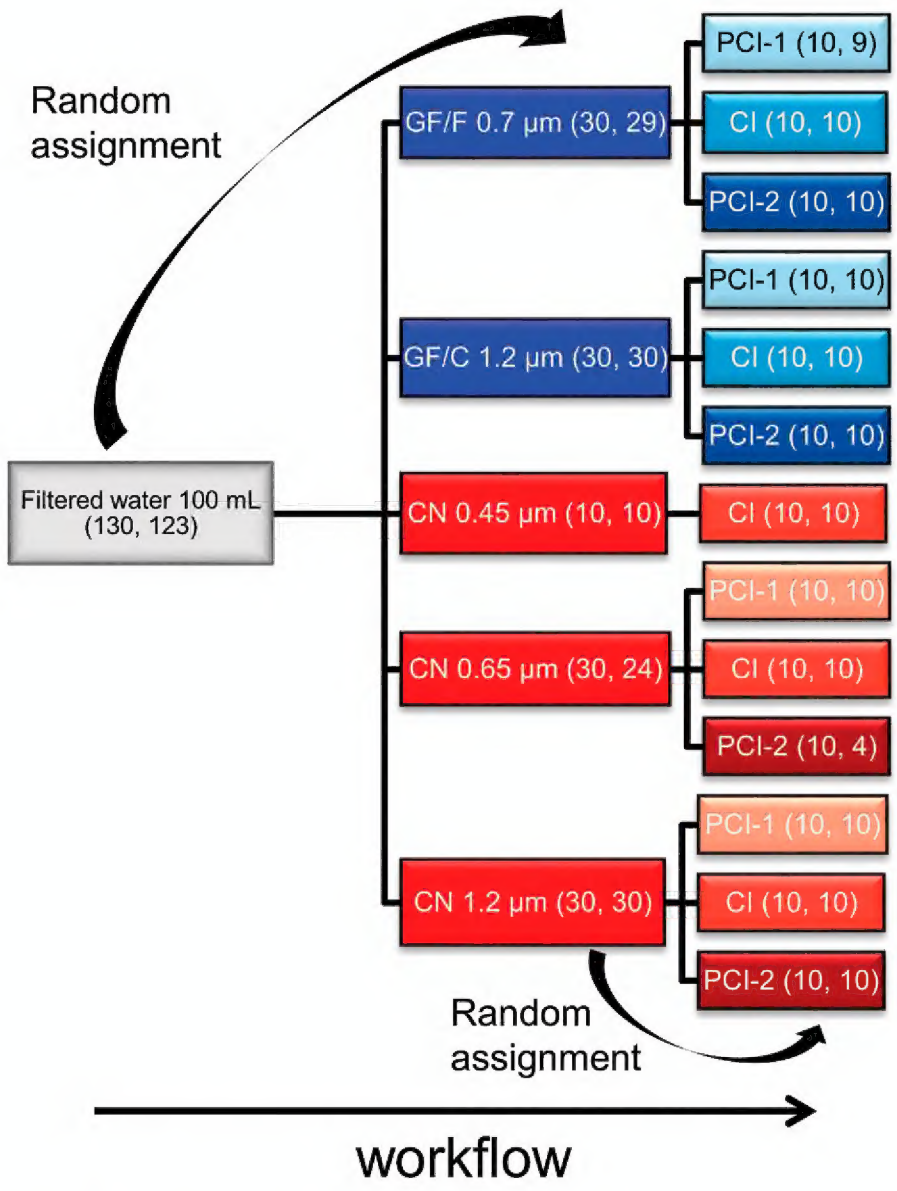

Figure 1. Experimental design used to test for an effect of filter material (Cellulose Nitrate, CN; Glass Fibre, GF) pore size and extraction method (PCI-1, PCI-2 and CI, see methods section for their description) on the detection eukaryotic DNA yield and the number of Operational Taxonomic Units (OTUs) estimated from seawater samples from a commercial port. Numbers in parentheses are the number of replicates with the target number first followed by the final number included in statistical analyses. Colours indicate experimental treatments where filter material is in blue (GF) or red (CN) and extraction methods are in variations of blue or red depending on filter material.

wave exposure by the natural harbour created by the Solent Estuary and the Isle of Wight. A modelling approach to identify invasion risk categorised the Solent as one of the highest risk areas for marine species invasions (Pearce et al. 2012). Furthermore, surveys of Southampton marinas have found many of the most common NIS in the UK, making it an ideal location to test whether eDNA metabarcoding methods are sensitive enough to detect NIS in commercial ports (Bishop et al. 2015).

Sample collection

To allow for broad taxonomic richness in the water used for the experiment, thirteen 1 litre seawater samples were collected from Empress Dock at two sites: seven from the harbour master's pontoon (50°53'24.7128"N, 1°23'47.2704"W) in the southwest corner and six from the National Oceanography Centre pontoon (50°53'28.2732"N, 1°23'38.0724"W) in the northeast corner. The 1 litre Nalgene™ plastic bottles were decontaminated with a 4% bleach solution (0.58% NaOCl) and rinsed using sterile Milli-Q water. Plastic bottles were then used to collect water from the top 0.5 m of the water column, first being rinsed out twice in the water and, finally, collecting the 1 litre water sample with the third fill. After collection, samples were immediately transported to the laboratory directly near the dock where they were refrigerated at 4 °C. Samples were refrigerated for 12.5 hours until filtration could begin.

Sample filtration

Prior to filtration, the thirteen 1 litre seawater samples were poured into a single container. Experimental replicates were created by taking 100 ml volumes of seawater from the single container (Figure 1). The single container of seawater was continuously mixed to minimise sedimentation of particulates and to ensure homogeneity across experimental replicates. Each 100 ml experimental replicate was randomly assigned to a filter type (Figure 1). To compare pore size and filter material, two pore sizes (0.7 µm and 1.2 µm) of Glass Microfibre Filters (GF) and three pore sizes (0.45 µm, 0.65 µm and 1.2 µm) of Cellulose Nitrate Filters (CN) were tested: GF 0.7 µm (Whatman), Grade GF 1.2 µm (Whatman), CN 0.45 µm (GE Healthcare), CN 0.65 µm (GE Healthcare) and CN 1.2 µm (GE Healthcare). All filter types had 47 mm diameters. Filters where chosen because they represent commonly used material and pore sizes. Pore sizes between GF 0.7 µm and CN 0.65 µm were considered equivalent pore sizes for statistical analysis of the experiment. Filtration occurred on a three-piece stainless-steel manifold with the capacity to processes three experimental replicates in parallel.

After each filtration, filters was folded and placed into a 2 ml microcentrifuge tube filled with 700 µl of Longmire's tissue lysis buffer (Longmire et al. 1997), ensuring that the filter was completely submerged and stored at -20 °C. The filters in Longmire's tissue lysis buffer were shipped to the University of Notre Dame at ambient temperature for processing. Longmire's buffer, as it is commonly called, has been shown to be effective for storage of filtered eDNA samples at room temperature (Renshaw et al. 2015). Use of a lysis buffer for preservation of eDNA samples is favourable because it has a pH suitable for stabilising DNA and additives capable of deactivating nucleases.

A rigorous cleaning protocol, before and during the procedure, ensured minimal cross-contamination between filters. All equipment was decontaminated before the procedure using a 4% bleach solution (0.58% NaOCl). The three-piece stainless-steel manifold and working area were first wiped down with the 4% bleach solution and then rinsed with distilled water and dried. New gloves and forceps were used for every experimental replicate and the decontamination procedure was repeated between replicates.

DNA extraction methods

Experimental replicates were randomly assigned by filter type to an extraction method as indicated in Figure 1. Three DNA extraction protocols were tested. Two consisted of phenol-chloroform-isoamyl protocols (PCI-1, PCI-2) and the third protocol used a chloroform-isoamyl protocol (CI). The exact step-by-step protocols that were followed are provided in Suppl. material 1. Protocol PCI-1 is a modified protocol that was developed by combining aspects of both the PCI-2 and the CI protocols. The PCI-1 protocol was previously used by Lacoursière-Roussel et al. (2018) and the main difference from either of the other two protocols is that it utilised QIA shredder filters (Qiagen, Germany), which have been reported to assist with performance of extracting DNA from environmental samples (Goldberg et al. 2013). Protocol PCI-2, which was previously used in comparison to commercial extraction kits and was shown to perform well when targeting both eukaryotes and bacteria (Deiner et al. 2015), was used here unmodified except to account for the difference in filter size used by increasing the volume of tissue lysis buffer added. Protocol CI was also previously used in comparison to kit extractions and is an inexpensive alternative that performed as well as other extraction methods (Renshaw et al. 2015). The CI protocol used in this study was unmodified. The CI protocol omits the chemical phenol, which is a toxic substance requiring extra safety precautions and added processing time and thereby considered a safer and less time-consuming option. DNA extractions were eluted in 75 µl of water or buffer as indicated in Suppl. material 1.

To remove potential inhibitors, extracted DNA was treated with the OneStep™ PCR Inhibitor Removal Kit (Zymo Research, Irvine, California, USA). Each replicate was quantified with 2 µl of sample using Qubit dsDNA High-Sensitivity Kit and Qubit 2.0 Fluorometer (Life Technologies, Grand Island, NY).

Library preparation methods

Our 18S ribosomal RNA gene primer set contained two sequences: Illumina adapter sequence and gene-specific target sequence. The Illumina adapter sequence provided the requisite flanking region for subsequent dual-indexing PCR and sequencing platform compatibility. The gene-specific sequence targeted a 378 bp fragment of the V4 variable region of the Eukaryote 18S ribosomal RNA gene. Hadziavdic et al. (2014) demonstrated *in silico* that the 18S_574F and 18S_952R primer set was the most optimal primer pair to cover eukaryote taxa while efficiently excluding prokaryotes. Together, these primers allowed for the generation of a eukaryotic 18S ribosomal RNA gene library compatible with Illumina sequencing platforms.

PCR-1

The library preparation workflow is illustrated in Figure 2. The eukaryote 18S ribosomal RNA gene V4 variable region (378 bp) was amplified from extracted DNA by high fidelity PCR using iProof PCR reagents (Bio Rad Laboratories, Hercules, CA), and our 18S ribosomal RNA gene primer set in a 0.2 ml PCR tubes with individually attached cap. The 50 µl-reaction contained 5.0 µl of extracted DNA (regardless of concentration), 27.0 µl of PCR-grade water, 10.0 µl of 5× HiFi Buffer (Bio Rad Laboratories, Hercules, CA), 1.5 µl of 50 mM magnesium chloride (Bio Rad Laboratories, Hercules, CA), 1.0 µl of 10 mM dNTP mix, (G-Bioscience, St. Louis, MO), 2.5 µl of 10 µM 18S_574F primer (Integrated DNA Technologies, Coralville, IA), 2.5 µl of 10µM 18S_952R primer (Integrated DNA Technologies, Coralville, IA) and 0.5 µl of 2 U/µl iProof HiFi DNA Polymerase (Bio Rad Laboratories, Hercules, CA). PCR reactions were cycled in a Master Pro S (Eppendorf, Westbury, NY) with the following conditions: initial denaturation at 98.0 °C for 2 minutes, followed by 25 cycles of denaturation at 98.0 °C for 10 seconds, annealing at 55.0 °C for 20 seconds and primer extension at 72.0 °C for 30 seconds, with a final extension at 72 °C for 10 minutes. Reactions were purified with a 0.8× bead:sample volume ratio (40.0 µl) of Agencourt Ampure XP Beads (Beckman Coulter, Webster, TX) and separated with a DynaMag-2 magnet rack (Life Technologies, Grand Island, NY). Ampure XP Beads were resuspended in 52.5 µl PCR-grade water to elute PCR products. After clearing the solution of Ampure XP Beads with the DynaMag-2 magnet, only 50.0 µl was recovered to avoid Ampure XP Bead carry-over. PCR product concentration was determined with 2.0 µl of sample using Qubit dsDNA High-Sensitivity Kit and Qubit 2.0 Fluorometer (Life Technologies, Grand Island, NY).

PCR-2

The Illumina adapter and dual-index barcodes (i7 and i5 index primers) were added to the PCR-1 products by high fidelity PCR using iProof PCR reagents (Bio Rad Laboratories, Hercules, CA) in a 0.2 ml PCR tube with individually attached cap. The 50 µl-reaction contained 5.0 µl of PCR-1 Template (regardless of concentration), 22.0 µl of PCR-grade water, 10.0 µl of 5× HiFi Buffer (Bio Rad Laboratories, Hercules, CA), 1.5 µl of 50 mM magnesium chloride (Bio Rad Laboratories, Hercules, CA), 1.0 µl of 10 mM dNTP mix, (G-Bioscience, St. Louis, MO), 5.0 µl of 10 µM i7 Index primer (Integrated DNA Technologies, Coralville, IA), 5.0 µl of 10 µM i5 Index primer (Integrated DNA Technologies, Coralville, IA) and 0.5 µl of 2 U/µl iProof HiFi DNA Polymerase (Bio Rad Laboratories, Hercules, CA). PCR reactions were cycled in a Master Pro S (Eppendorf, Westbury, NY) with the following conditions: initial denaturation at 98.0 °C for 2 minutes, followed by 8 cycles of denaturation at 98.0 °C for 10 seconds, annealing at 60.0 °C for 20 seconds and extension at 72.0 °C for 30 seconds, with a final extension at 72.0 °C for 10 minutes. Reactions were purified with a 0.8× bead:sample volume ratio (40.0 µl) of Agencourt

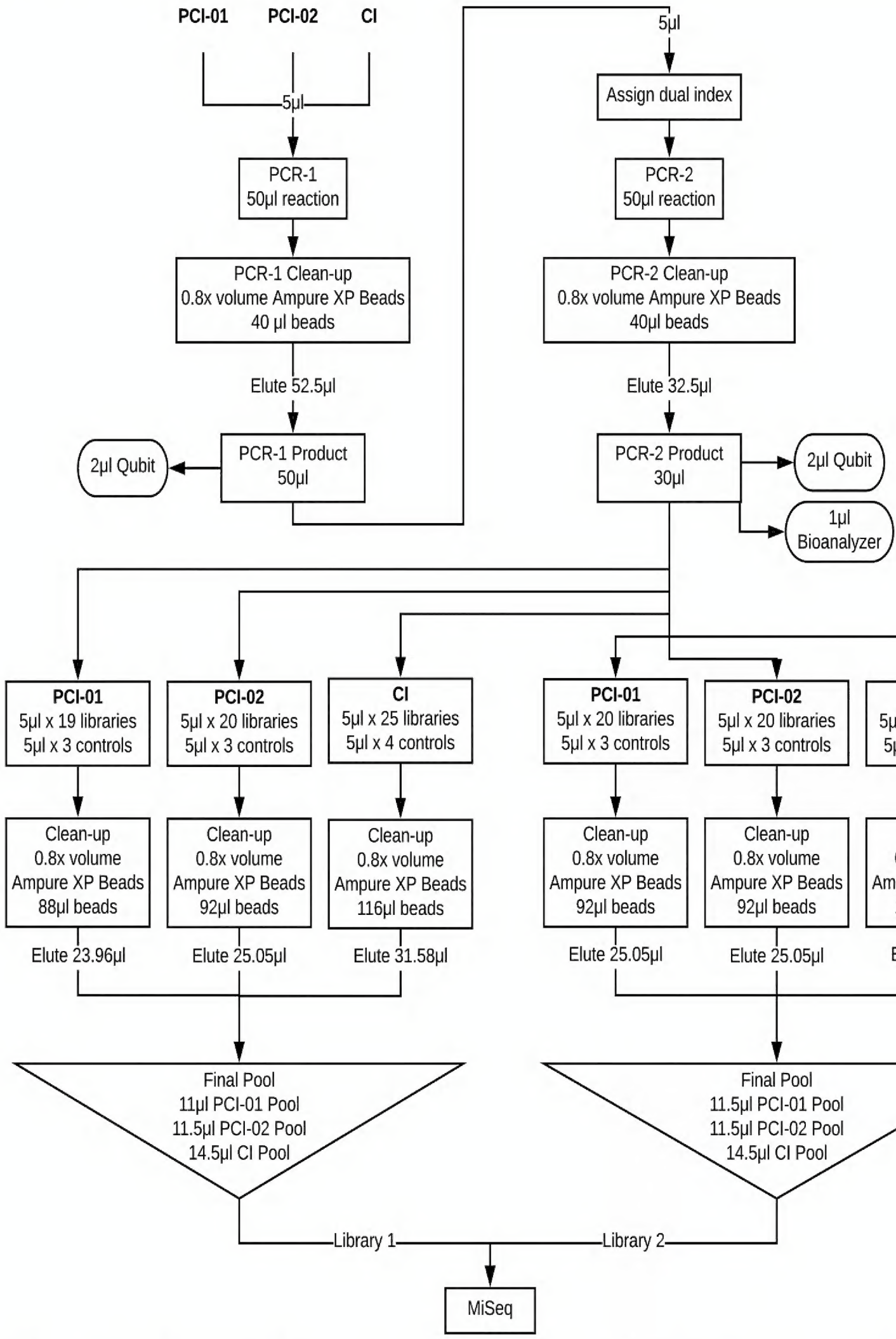

Figure 2. Library preparation workflow illustrating pooling by volume rather than by normalising to avoid removing the effect of the experimental treatments.

Ampure XP Beads (Beckman Coulter, Webster, TX) and separated with a DynaMag-2 magnet (Life Technologies, Grand Island, NY). Ampure XP Beads were resuspended in 32.5 µl PCR-grade water to elute the dual-indexed library. After clearing the solution of Ampure XP Beads with the DynaMag-2 magnet, only 30.0 µl was recovered to avoid Ampure XP Bead carryover. Library concentration was determined with 2.0 µl of sample using Qubit dsDNA High-Sensitivity Kit and Qubit 2.0 Fluorometer (Life Technologies, Grand Island, NY).

No-template controls

No-template controls were introduced at multiple steps in the workflow (DNA extraction, eukaryotic 18S ribosomal RNA gene PCR-1 and PCR-2). No-template controls followed the same procedure as experimental replicates. In place of extracted DNA, 5.0 µl of PCR-grade water was used as the template for PCR-1 and PCR-2. A total of ten no-template controls were generated to detect contamination during seawater sampling and preparation of libraries.

Library Validation

Due to the project's scale, 31 of 129 dual-indexed libraries (~24%), derived from experimental replicates, were randomly chosen for Bioanalyzer analysis to control overall cost of validation. To verify the final library fragment size and diagnose primer contamination, 1.0 µl of dual-indexed library was analysed using Bioanalyzer 2100 instrument and Bioanalyzer DNA 7500 kit (Agilent Technologies, Santa Clara, CA). All no-template controls (1.0 µl) were analysed using Bioanalyzer DNA High-Sensitivity kit (Agilent Technologies, Santa Clara, CA).

Multiplexing Libraries

Due to the number of samples in the study, the libraries where split across two MiSeq runs. To avoid a run effect, libraries were randomly assigned to one of two pools by sub-group (explained below), extraction method and filter material. All ten no-template control libraries were added to both pools to test for consistency between runs. As PCR-2 DNA concentrations from each filter material were the response variables, the volume of template added to the eukaryotic 18S ribosomal RNA gene PCR and dual-index PCR was held constant without normalising the template concentration as this would remove the effect of our experiment (Fig. 2). Furthermore, we multiplexed equal volumes of library or no-template control by group and pool assignment (Fig. 2). Specifically, 5.0 µl of library (regardless of concentration) and 5.0 µl of no-template control were combined by group. This created six (6) sub-pools: PCI-1 Pool-1, PCI-1 Pool-2, PCI-2 Pool-1, PCI-2 Pool-2, CI Pool-1 and CI Pool-2. To remove primer contamination, each sub-pool was purified with a 0.8× bead:sample volume ratio of Agencourt Ampure XP Beads (Beckman Coulter, Webster, TX) and separated with a DynaMag-2 magnet (Life Technologies, Grand Island, NY). Ampure XP Bead volume was calculated by multiplying 0.8 and the total volume of the sub-pool. Ampure XP Beads were resuspended with 1.089 µl PCR-grade water per library and no-template control in the sub-pool. This volume was chosen based on the desire for an elution volume of about 32 µl to effectively concentrate the pooled libraries by 4 to 5 times. In practice, 32 µl would be divided by the sum of the number of libraries and no-template controls in the pool to determine the elution volume per library (Fig. 2). (Example: If a sub-pool with 25 libraries and 4 no-template controls had a total volume of 145 µl, then116 µl of Ampure XP Beads would be added. Ampure XP Beads would be resuspended with 31.58 µl of PCR-grade water and then 29.0 µl recovered after bead separation.). After clearing the solution of Ampure XP Beads with the DynaMag-2 magnet, 2.5 µl less than the elution volume was recovered to avoid Ampure XP Bead carry-over. Finally, 0.5 µl per library and no-template control of the concentrated sub-pools were combine by final pool assignment. (Example: Subpool Group M with 25 libraries and 4 no-template controls contributed 14.5 µl to the final pool while Sub-pool Group K with 20 libraries and 3 no-template controls contributed 11.5 µl to the final pool).

Final Library Validation and Illumina Sequencing

The Genomics Core Facility (University of Notre Dame, Notre Dame, IN) validated the final library pools using a combination of Qubit dsDNA High-Sensitivity Kit and Qubit 2.0 Fluorometer (Life Technologies, Grand Island, NY), Bioanalyzer 2100 instrument and Bioanalyzer DNA 7500 Kit or Bioanalyzer DNA High-Sensitivity Kit (Agilent Technologies, Santa Clara, CA) and Kapa Illumina Library Quantification qPCR assay (Kapa Biosystems, Wilmington, MA). Final molar concentration of the final library pool was based on the average of the values determined by Qubit and qPCR analysis. For each of the final library pools, a solution containing 6 pM denatured library and 3 pM denatured PhiX Control v3 (Illumina, Inc., San Diego, CA) was sequenced on Illumina MiSeq Sequencer operating MiSeq Control Software v2.5 with MiSeq flowcell (v3) and MiSeq Reagent Kit (v3). Sequencing format was 300 cycles for read 1, 8 cycles for index 1 read, 8 cycles for index 2 read and 300 cycles for read 2. Each library pool was run on a different MiSeq flow cell. MiSeq Reporter v2.5 Real Time Analysis (RTA) v1.18.54 performed base calling. Illumina Bcl2fastq v2.18 demultiplexed the RTA output and converted the data from bcl to fastq format. Finally, BaseSpace Broker v2.1 reported the files to BaseSpace Sequencing Hub.

Lastly, one set of samples was unfortunately mixed during the second stage of DNA extraction. While they were fully processed and contributed to the total raw reads observed, we removed them from any statistical analysis due to not being able to differentiate which replicate they were. This error reduced the number of experimental replicates for the treatment of cellulose nitrate filters with a pore size of 0.65 µm (CN 0.65 µm) extracted with the PCI-2 protocol from ten to four. We also lost one sample during DNA extraction from the treatment with the filter material of glass fibre with a pore size of 0.7 (GF 0.7 µm) extracted with PCI-1. The replicate could not be processed for sequencing and thus reduced the number of replicates for this treatment from ten to nine.

Bioinformatic analysis

Sequencing adapters were removed and reads were quality filtered using Trimmomatic v0.32 (Bolger et al. 2014). An in-house Perl script was used to check for 100% match to both forward and reverse primers (Github: https://github.com/pfrender-laboratory/epps) and non-matching reads were discarded. Paired end reads were then merged using the -fastq_mergepairs command in USEARCH v8.1.1861(Edgar 2013) and singleton merged sequences (defined as a merged sequence observed only once across all replicates) were discarded. All replicates were pooled for OTU generation and OTU clustering was performed

using the -cluster_otus command in USEARCH with two clustering thresholds of 97% and 95% similarity. Potential chimera sequences were also filtered by -cluster_otus command during the OTU clustering process as suggested by USEARCH. Centroid sequences from each OTU were taxonomically annotated using the Species Assignment Package v1.9.4 (Munch et al. 2008) against the NCBI nr database. OTUs were counted as metazoan if they received an assignment to the phylum level or lower. Specific parameters for each filtration step are shown in Suppl. material 2.

To explore the effect of method choice on a globally relevant NIS, we determined the detection of *Styela clava* in different technical replicates based on species assignments of OTUs. *Styela clava*, a solitary ascidian, is native to the NW Pacific. Over the last 100 years, it has become common in fouling communities in harbours and ports across the globe, including Europe, Australasia and North America (Goldstien et al. 2011). In the United Kingdom, it was first discovered in 1953 in Plymouth and then described as *Styela mammiculata* by Carlisle (1954). This introduction was attributed to the transport via warships after the end of the Korean war (Holmes 1968). It was first recorded in Southampton water in 1958 reaching population densities of 200 adults per square metre (Holmes 1968). Since then, the contemporary presence of *Styela clava* in Southampton water is well documented (Bishop et al. 2015). In each replicate, detection was considered positive if a single read or greater was assigned to *Styela clava*.

Statistical analysis

All statistical analyses were performed using the programme R, version 3.3.1.1 (R Core Team 2016). R scripts and required data files are provided on Figshare (https://doi.org/10.6084/m9.figshare.c.4191167.v1). Least squares regression models were used to test the additive effect of filter material (GF or CN), extraction type (CI, PCI-1 and PCI-2), pore size (0.5 µm, 0.7 µm and 1.2 µm) and the nested effect of filter material on pore size on PCR-2 DNA concentrations and OTU richness (both eukaryotic and metazoans only). As mentioned above, GF 0.7 µm and CN 0.65 µm were considered equivalent pore sizes for statistical analysis of the experiment. Richness of an experimental replicate was calculated by summing all OTUs with a single read or more. Due to the possible violation of homogeneity with pore size, resulting from the imbalance in factor levels (0.5 µm = 10, 0.7 µm = 53, 1.2 µm = 60), we also tested the models after removing the smallest pore size (0.5 µm) from the dataset. In addition to the two response variables of PCR-2 DNA concentrations and OTU richness, the variables of DNA concentration after PCR-1, the number of raw merged reads, number of merged reads after quality filtering, the number of OTUs clustered at 95% and the subset of metazoan OTUs clustered at 95% were tested for a correlation to establish an association between each step in the laboratory workflow and the outcome of methods' choices. To evaluate the detectability of the NIS *Styela clava* across the experimental treatments, reads from the five OTUs assigned to *Styela clava* were pooled (Suppl. material 3) and detection was indicated by a single read or more. The effect of experimental treatments on detection was modelled using the same least squares regression model as above.

Results

Sequencing and OTUs

Raw fastq files have been deposited in the Short Read Archive (PRJNA395904). The two MiSeq runs produced a total of 45,600,412 raw reads for the experimental replicates and 41,522 reads for the no-template controls. After merging paired reads and quality filtering, a total of 16,144,713 reads remained for the experimental replicates and 837 reads for the no-template controls. Experimental replicates had an average of 125,153 ± 111,631 reads and no-template controls had an average of 42 ± 25 reads.

The number of unique reads summed across both runs was 996,242. After removal of singletons, 260,711 unique reads remained. Clustering at 97% and 95% resulted in 4,554 and 3,634 OTUs, respectively (Suppl. material 3). Based on taxonomic assignment of OTUs clustered at 97% and 95%, 2,112 and 1,519 were identified to at least the taxonomic level of phylum and represented 123 and 102 metazoan eukaryotes, respectively. Using clustering at 97% as the example, the no-template controls had reads assigned to 61 of the OTUs (Suppl. material 4). Of these, 28 OTUs had only one read assigned to a no-template control inclusive of both MiSeq runs. When comparing the other 33 OTUs, only two were consistently detected between the two MiSeq runs. As so few OTUs were consistently detected between runs and nearly half were represented by one read, we did not consider this evidence of contamination due to preparation error in the sampling or library preparation workflow, but rather evidence of the known tag-jumping or cross-talk errors generated on Illumina instruments (Schnell et al. 2015; Edgar 2016). As this error results in so few reads, we considered this error random and did not apply a correction factor to the data.

DNA concentration

DNA concentration varied as a function of filter material, pore size and extraction method (Fig. 3a). Using PCR-2 DNA concentrations as the response variable for an estimate of DNA concentrations of eukaryotic environmental DNA (Fig.3b), a significant effect of extraction type ($P < 0.001$), pore size ($P < 0.001$), filter material ($P < 0.001$) and a nest effect of pore size on filter material ($P = 0.035$) were found (Table 1, Fig. 4a). Specifically, DNA concentration after PCR-2 ranged from 0.08 to 45.70 ng/µl (mean = 18.427, s.d. = 15.770) (Fig. 4a). Chloroform extractions yielded PCR-2 DNA concentrations from 0.08 to 30.6 ng/µl (mean = 4.314, s.d. = 6.429), with PCI-1 extractions pro-

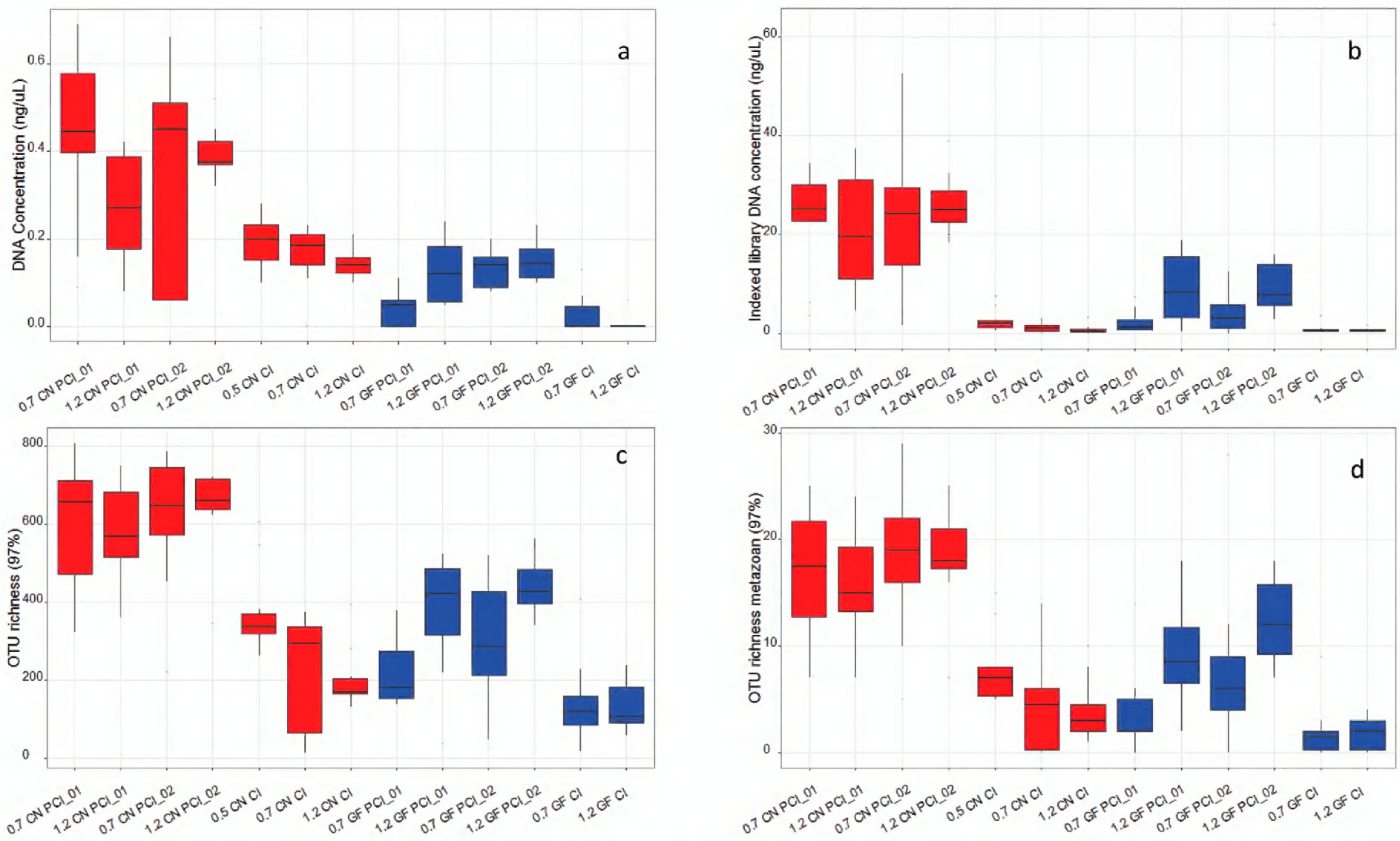

Figure 3. Observed DNA concentrations or number of Operational Taxonomic Units by experimental treatment. Observed DNA concentrations after DNA extraction (**a**) indexed PCR (PCR-2) (**b**) and the number of estimated OTUs per experimental treatment clustered at 97% for all eukaryotes and (**c**) and metazoans (**d**). The upper and lower whiskers indicate the minimum and maximum point within 1.5 times the Interquartile Range extended from the 25th and 75th percentile, respectively. Colours indicate filter material where red is Cellulose Nitrate (CN) and blue is Glass Fibre (GF). The three extraction methods are abbreviated as in Figure 1.

ducing 0.18 to 45.7 ng/µl (mean = 26.414, s.d. = 12.861) and PCI-2 extractions producing 0.16 to 45.5 ng/µl (mean = 30.022, s.d. = 12.323). DNA concentrations, after PCR-2, varied with corresponding pores sizes as well as with samples filtered with 0.5 µm filters showing 2.78 to 30.6 ng/µl (mean = 11.173, s.d. = 9.780), 0.7 µm resulting in 0.08 to 45.7 µm (mean = 15.849, s.d. = 15.453) and 1.2 µm filters resulting in 0.15 to 45.5 ng/µl (mean = 21.914, s.d. = 16.184). Samples processed with CN filters had PCR-2 DNA concentrations of 0.08 to 45.7 ng/µl (mean = 22.021, s.d. = 16.682) and samples processed with GF filters had PCR-2 DNA concentrations of 0.09 to 42.6 ng/µl (mean = 14.400, s.d. = 13.733). Models, from which we removed the smallest pore size to assess possible type I errors, showed significant differences across extraction method ($P < 0.001$), pore size ($P = 0.004$) and filter material ($P < 0.001$) and between the interaction of pore size nested within filter material ($P = 0.033$) (Suppl. material 5).

OTU richness

OTU detection for eukaryotes and the subset of metazoan eukaryotes varied with method choice (Figure 3c, d). DNA extraction concentrations and PCR-2 DNA concentrations were correlated with OTU richness estimated for all eukaryotes and the subset of metazoan eukaryotes (Fig. 5). Significant effects were observed for the number of OTUs (clustered at 97% similarity) detected across ex-

Table 1. Statistical results from the least squares regression models with response variables of PCR-2 DNA concentrations, OTUs with 97% sequence similarity and *Styela clava* detection. Provided for each explanatory variable (extraction type, filter material, pore size and pore size nest in filter material) are the degrees of freedom (DF), Sum of squares (Sum sq), Mean squares (Mean sq), F value and p-value (P-value).

	DF	Sum Sq	Mean Sq	F value	P value
Response: PCR-2 DNA Concentration					
Extraction type	2	17018.2	8509.1	111	< 0.001
Pore size	2	1231.1	615.5	8.06	< 0.001
Filter material	1	2883.3	2883.3	37.8	< 0.001
Pore size / Filter material	1	347.8	347.8	4.55	0.035
Residuals	116	8859.8	76.4		
Response: OTU 97					
Extraction type	2	1980701	990351	55	< 0.001
Pore size	2	378264	189132	10.5	< 0.001
Filter material	1	1152720	1152720	64	< 0.001
Pore size / Filter material	1	66101	66101	3.67	0.058
Residuals	116	2090261	18019		
Response: *Styela clava* detection					
Extraction type	2	4.148	2.074	10.595	<0.001
Pore size	1	0.177	0.177	0.905	0.343
Filter material	1	0.587	0.587	2.998	0.086
Pore size / Filter material	1	0.007	0.007	0.033	0.855
Residuals	123	24.074	0.196		

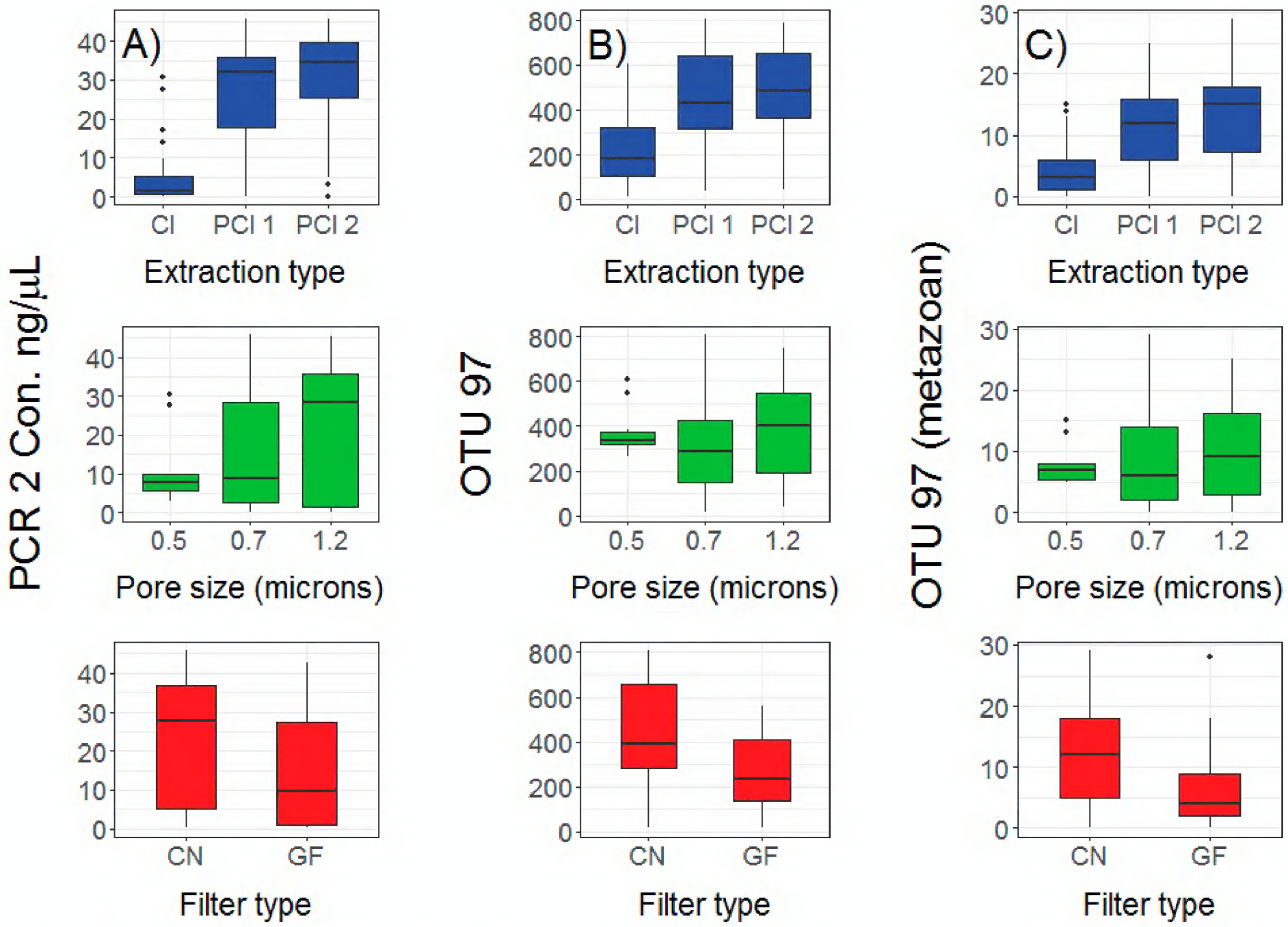

Figure 4. Observed DNA concentrations or number of Operational Taxonomic Units by experimental factor. Mean PCR-2 DNA concentrations (indexed libraries) (y-axis) for each set of explanatory variables including extraction type (**A**), filter pore size (**B**) and filter material (**C**). Extraction types include chloroform (Cl), phenol chloroform 1 (PCl 1) and phenol chloroform 2 (PCl 2) extractions. Filter materials included Cellulose Nitrate (CN) and Glass Fibre (GF). The upper and lower whiskers indicate the minimum and maximum point within 1.5 times the Interquartile Range extended from the 25th and 75th percentile, respectively.

traction types (P <0.001), pore size (P = 0.010) and filter material (P <0.001) (Fig. 4b, c). However, a non-significant interaction effect of pore size nested in filter material (P = 0.058) was detected (Table 1, Fig. 4b, c). Models, from which the smallest pore size was removed to assess possible type I errors, showed comparable results in that OTU detection and significant differences amongst extraction types (P <0.001), pore size (P = 0.010) and filter material (P <0.001), but there was a non-significant effect of pore size nested in filter material (P = 0.056) and a non-significant effect of pore size (P=0.061) (Fig. 4b) (Suppl. material 5). Results were similar for the subset of OTUs that were assigned to the metazoan phylum and clustered at 95% similarity (Fig. 4c, Fig. 5). Furthermore, each step of the workflow from DNA extraction to library preparation to bioinformatic choices for clustering threshold showed strong positive correlations (Fig. 5).

Based on the taxonomic assignment of OTUs, *Styela clava* was detected in 45 of 73 (62%) PCI replicates compared to 10 of 50 (20%) CI extraction replicates (Suppl. material 3). A significant effect (P = 0.001) of extraction method was found; the CI method was less effective than both phenol-based extractions (Table 2, Fig. 6).

Discussion

The eDNA metabarcoding methods experiment findings show that filter material, pore size and extraction method affects the yield in marine eukaryotic eDNA. Additionally, the filter material and extraction method, but not pore size, influence the estimated richness of OTUs detected from seawater using eDNA metabarcoding. Specifically, as many as three times the number of OTUs could be detected by adjusting the filter material and extraction method, indicating that some materials and methods combinations have high false negative detection rates. While no other studies have tested for the effect of all three variables tested here, these results corroborate experiments in freshwater systems where cellulose nitrate filters yielded a higher DNA concentration compared to other filter materials (Hinlo et al. 2017; Lacoursière-Roussel et al. 2016). Specifically, for the commercial port seawater sampled here, maximum OTU richness was detected with cellulose nitrate filters and phenol-chloroform-isoamyl extraction. In addition, the largest pore size tested, 1.2 µm, performed indistinguishably from the 0.65 µm and

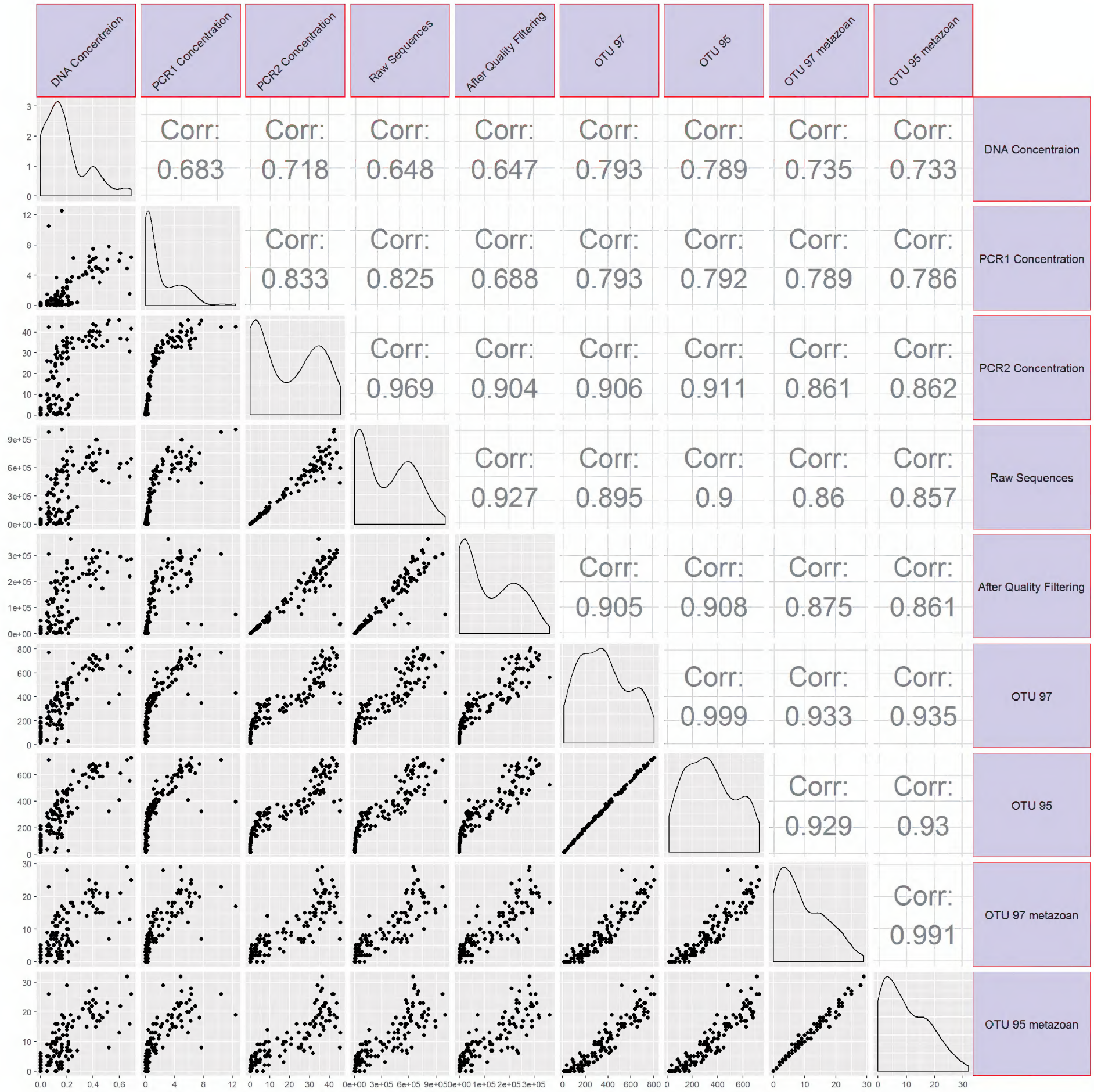

Figure 5. Pairwise correlations between non-independent response variables corresponding to quantified values taken during the analyses. Lower triangle consists of scatterplots with individual sample comparisons shown. The upper triangle shows the R correlation value corresponding with the lower triangle. The diagonal shows the distribution curve for the variable indicated in the corresponding column.

0.7 µm pore sizes (Figure 3). As larger pore sizes allow for shorter filtration times (Li et al. 2018), if we were to implement an eDNA sampling regime for this environment, these results suggest that the most effective and efficient materials and methods' combination would be a 1.2 µm cellulose nitrate filter with the phenol-chloroform-isoamyl extraction protocol (PCI-2).

An important observation from this study is that all materials and methods used produced biodiversity information, even when DNA yields were below detection limits (e.g. Figure 3a, CI extraction). However, by adjusting the filter material and extraction method, we could reduce the false negative rate by as many as two orders of magnitude. Thus, unless methods are compared, it is hard to determine their OTU detection limits. Increasing OTU detection rates for species in commercial ports is valuable for monitoring of rare or endangered species or where novel species invasions require elevated levels of biosecurity. Indeed, marine eDNA studies have already begun to detect both non-indigenous (Borrell et al. 2017; Simpson et al. 2017) and rare species (Weltz et al. 2017). For example, the invasive ascidian *Styela clava*, which was observed as present during sampling of the studied site, was detected 40% more often in phenol-chloroform-isoamyl extraction replicates and achieved a detection rate as high as 80% compared to a 50% detection rate for CI extraction replicates. Our results, based on both the detection of *Styela clava* and the increase of OTUs detected,

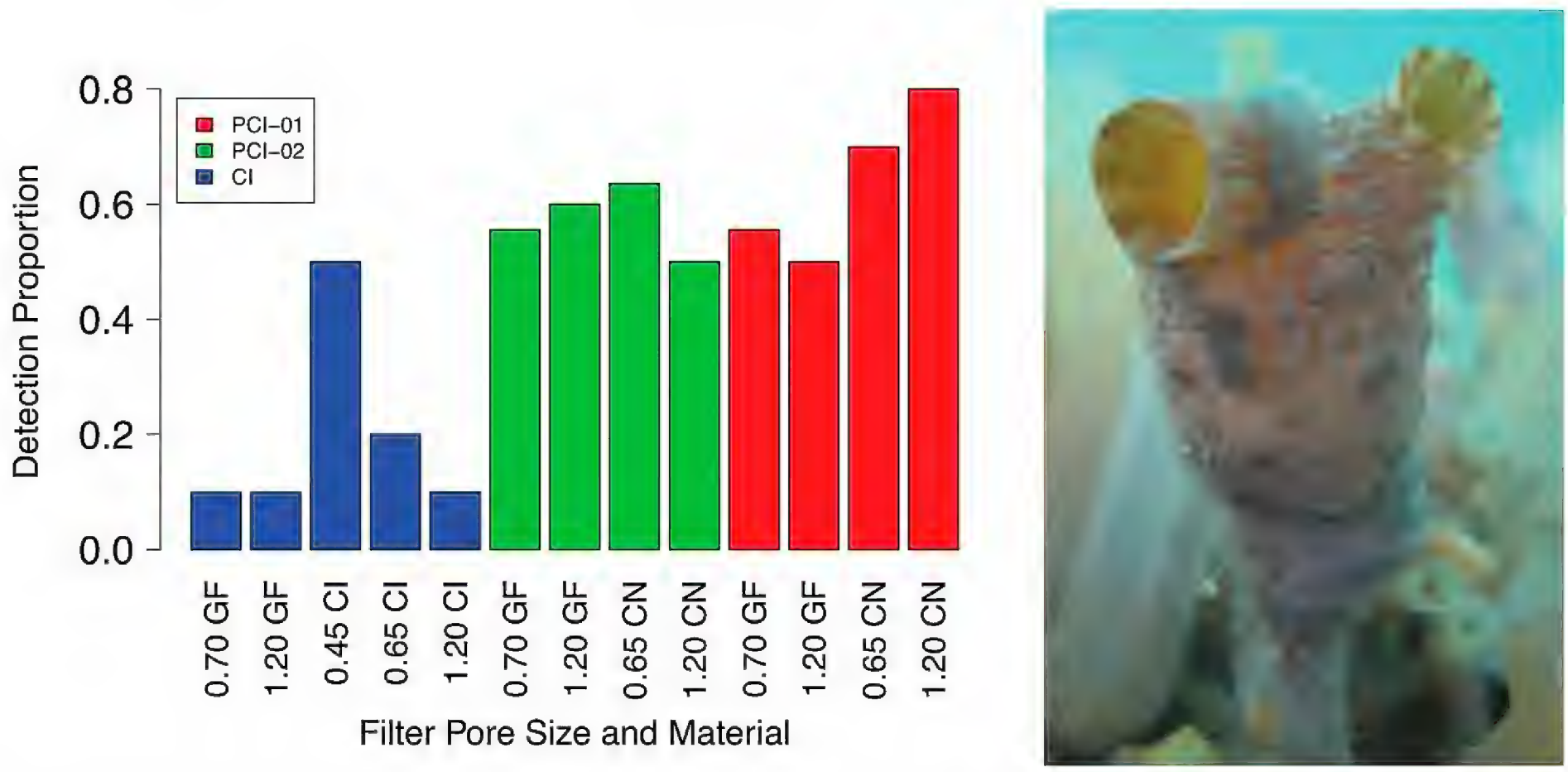

Figure 6. Detection rate of *Styela clava* by extraction method. The proportion of positive detections for *Styela clava* across experimental treatments (left). Photographic image of *Styela clava* individual from Chichester Harbour, United Kingdom (Right). Colours are indicating the three different extraction methods which are abbreviated the same as in Figure 1.

demonstrate that it is essential to understand the detection rate of material and method choices before making an assessment about species presence or absence.

The logical conclusion from first principles, which infers that false negative detections decrease as DNA yield from a sample increases, suggests that the DNA extraction is an obvious step in the process to optimise. The experimental test here demonstrated that the phenol-chloroform-isoamyl DNA extraction methods produced a greater yield of DNA compared to the chloroform-isoamyl extraction method. The slight modifications between the two phenol-chloroform-isoamyl DNA extractions methods tested were not significantly different. Previous work in freshwater systems have also shown that a phenol-chloroform-isoamyl DNA extraction method resulted in a higher yield of DNA compared to DNeasy Animal Tissue kits (Qiagen) (Deiner et al. 2015). However, these results were based on a single filter material, glass fibre filters. More recent work indicates, inclusive of this study, that there are interactions between filter material and extraction methods, indicating that both filter material and extraction interact to drive yields in DNA (Djurhuus et al. 2017). Thus, it may be possible to maximise eDNA yield by jointly optimising which extraction methods performs best on a filter material such that biodiversity detected is equivalent between any chosen materials and methods. Phenol-chloroform-isoamyl methods perform best with cellulose nitrate filters compared to glass fibre filters in this study, and we recommend that future research should determine other combinations to optimise eDNA yield from other filter material and extraction combinations.

This study is the first to compare filter material and similar pore sizes between filter materials. However, including a statistical term where pore size was nested within filter material provide a better model fit to the data compared to a purely additive model. The non-independence of filter material with pore size and estimated richness is likely because filter pore sizes are not uniform in structure between different filter materials. In fact, glass fibre filters are described by the manufacturer as having a nominal pore size of 0.7 µm, but the size of any one pore can vary. For a micrograph of pore structures of different filter materials, see the Supplemental material from figure S2 of Turner et al. (2014). Thus, there may be synergistic physical or chemical interactions generated with pore size and different filter materials. Follow up work is needed to determine how the pore's structure and filter material interact with eDNA to yield differences in eDNA concentration and detection of OTU richness. Sequential filtering experiments in fresh (Turner et al. 2014) and salt water systems (Sassoubre et al. 2016) have shown that eDNA exists in a range of particle sizes. Therefore, it is expected that a smaller pore size would capture more eDNA with a trade-off of smaller pores requiring more effort to filter as the pores become blocked (Li et al. 2018). The trade-off between filtration time needed to attain a specific volume and the pore size interacting with the particulate matter in water samples clogging pores depends on the study system and question. We therefore recommend that pilot studies are conducted to determine which pore size is adequate for sampling eDNA given local environmental conditions and which results in an appropriate detection rate.

The choices made about the amount of replication in eDNA methodology experiments require more attention. Replication amongst eDNA methodology studies varies, with some including only three replicates per treatment (Djurhuus et al. 2017; Lacoursière-Roussel et al. 2016; Piggott 2016; Spens et al. 2017) and others choosing 4 to 15

replicates per treatment (Deiner et al. 2015; Eichmiller et al. 2016; Hinlo et al. 2017). However, none of these studies presented justification for the level of replication. A statistical power analysis was *a priori* used in this study to determine the number of replicates needed and extra precaution was taken to use twice as many to allow for the inadvertent loss of samples due to technical or procedural error. Reporting insignificant differences between methods without enough power to detect the effect can impact proper method choice, both in the study and for latter work built upon the inference. We recommend future methodological studies use a power analysis or other such methods to provide information about the experimental design such that non-significant differences can be interpreted as meaningful.

The current practice in eDNA metabarcoding studies is to normalise DNA concentration of prepared libraries from each sample before pooling and sequencing. The intention of which is to produce a dataset with equal sequencing depth per sample. The normalisation of read depth per library can also be achieved bioinformatically (Bista et al. 2017; Kelly et al. 2016). The practice of normalising DNA concentration across libraries is logical when the question being tested can be confounded by sequence depth. However, when the variable being tested is, in fact, the yield in DNA between samples, normalising DNA concentration at any step before sequencing would remove the effect from the experiment. Therefore, normalisation by volume rather than DNA concentration was used in this study (Fig. 2). Normalisation of DNA concentration between samples during library preparation at any step is not always clearly reported in the methodology; therefore, it is unclear how many eDNA metabarcoding method studies may have unintentionally decreased the likelihood of finding an effect related to yield of DNA from an environmental sample. Future work should examine the effect of normalisation during library preparation for method comparison studies, as it was not the goal here. By comparing normalised libraries to unnormalised, it may be possible to decrease the effect of low DNA yield from a set of method choices. This is identical to sequencing more DNA from an amplified extraction, but this alternative method choice assumes the original pool of environmental sequences is not biased by the initial low yield generated from the DNA extraction itself. Furthermore, the normalisation of samples may have a dramatic negative effect on the detected biodiversity in case where a single dominant species overwhelms the sample, as commonly found in oceans (jellyfish bloom, krill swarm, phytoplankton bloom). This effect has been documented by Deagle et al. (2018) and future studies should explore the effects of normalisation on species detectability in cases where the total DNA diversity of the ecosystem is 'swamped' by a small number of species.

Conclusions

While the advice and methods for detecting false positives have matured and researchers implement these practices (Goldberg et al. 2016), there is still considerable work ahead to understand what generates false negative detections. A clear recommendation to maximise detection of OTU richness emerging from this study is that cellulose nitrate filters combined with a phenol-chloroform-isoamyl extraction method yields more eDNA and detects a greater richness of OTUs than other method combinations. A strong positive association between DNA yield and OTU richness and all the steps in between was observed in this study (Fig. 5) and indicated that thresholds in library concentrations should be sought to help reduce false negative detections.

This study adds to the growing literature that false negative detections are a consequence of how the sample is processed in the laboratory. While methods will continue to be optimised and because false negative detections are often the result of a sampling problem, all steps where the sample of eDNA is biased can create false negative detections. Continued work is needed to identify the most crucial steps where this bias is introduced and future research on the laboratory methods should focus on optimisation of steps where the most gain in species detection is possible. These include, but are not limited to, the sequencing depth (Grey et al. 2018) and the total amount of DNA extracted and subsequently used in PCR (Mächler et al. 2015). Eventual standardisation of these methods when used in management settings is needed to produce actionable results. Nevertheless, the size of eDNA containing particles may differ amongst taxa and environments and other conditions affecting detection (e.g. the concentration of PCR inhibiting material), thus researchers and managers should consider that multiple methods may be needed for different locations. Efforts to develop standards are being made through the European DNAqua-net Cost Action (Leese et al. 2016) in collaboration with the European Committee for Standardisation (http://www.cen.eu).

Acknowledgements

This research was funded by the NSF Coastal SEES grant #1427157 (to DML, EKG). SB and LEH were supported by Natural Environment Research Council Grant #NE/L002531/1. LB and ALR were supported by Polar Knowledge Canada # NST-1617-0016B,

References

Barnes MA, Turner CR (2016) The ecology of environmental DNA and implications for conservation genetics. Conservation Genetics 17: 1–17. https://doi.org/10.1007/s10592-015-0775-4

Bishop JD, Wood CA, Yunnie AL, Griffiths CA (2015) Unheralded arrivals: non-native sessile invertebrates in marinas on the English coast. Aquatic Invasions 10. https://doi.org/10.3391/ai.2015.10.3.01

Bista I, Carvalho GR, Walsh K, Seymour M, Hajibabaei M, Lallias D, Christmas M, Creer S (2017) Annual time-series analysis of aque-

ous eDNA reveals ecologically relevant dynamics of lake ecosystem biodiversity. Nature Communications 8: 14087. https://doi.org/10.1038/ncomms14087

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30: 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Borrell YJ, Miralles L, Do Huu H, Mohammed-Geba K, Garcia-Vazquez E (2017) DNA in a bottle-Rapid metabarcoding survey for early alerts of invasive species in ports. PloS ONE 12: e0183347. https://doi.org/10.1371/journal.pone.0183347

Carlisle D (1954) *Styela mammiculata* n. sp., a new species of ascidian from the Plymouth area. Journal of the Marine Biological Association of the United Kingdom 33: 329–334. https://doi.org/10.1017/S0025315400008365

Deagle BE, Clarke LJ, Kitchener JA, Polanowski AM, Davidson AT (2018) Genetic monitoring of open ocean biodiversity: An evaluation of DNA metabarcoding for processing continuous plankton recorder samples. Molecular Ecology Resources 18: 391–406. https://doi.org/10.1111/1755-0998.12740

Deiner K, Bik HM, Mächler E, Seymour M, Lacoursière-Roussel A, Altermatt F, Creer S, Bista I, Lodge DM, de Vere N (2017) Environmental DNA metabarcoding: transforming how we survey animal and plant communities. Molecular Ecology 26: 5872–5895. https://doi.org/10.1111/mec.14350

Deiner K, Walser J-C, Mächler E, Altermatt F (2015) Choice of capture and extraction methods affect detection of freshwater biodiversity from environmental DNA. Biological Conservation 183: 53–63. https://doi.org/10.1016/j.biocon.2014.11.018

Díaz S, Demissew S, Carabias J, Joly C, Lonsdale M, Ash N, Larigauderie A, Adhikari JR, Arico S, Báldi A (2015) The IPBES Conceptual Framework-connecting nature and people. Current Opinion in Environmental Sustainability 14: 1–16. https://doi.org/10.1016/j.cosust.2014.11.002

Djurhuus A, Port J, Closek CJ, Yamahara KM, Romero-Maraccini O, Walz KR, Goldsmith DB, Michisaki R, Breitbart M, Boehm AB (2017) Evaluation of filtration and DNA extraction methods for environmental DNA biodiversity assessments across multiple trophic levels. Frontiers in Marine Science 4: 314. https://doi.org/10.3389/fmars.2017.00314

Edgar RC (2013) UPARSE: highly accurate OTU sequences from microbial amplicon reads. Nature Methods 10: 996. https://doi.org/10.1038/nmeth.2604

Eichmiller JJ, Miller LM, Sorensen PW (2016) Optimizing techniques to capture and extract environmental DNA for detection and quantification of fish. Molecular Ecology Resources 16: 56–68. https://doi.org/10.1111/1755-0998.12421

Evans NT, Li Y, Renshaw MA, Olds BP, Deiner K, Turner CR, Jerde CL, Lodge DM, Lamberti GA, Pfrender ME (2017) Fish community assessment with eDNA metabarcoding: effects of sampling design and bioinformatic filtering. Canadian Journal of Fisheries and Aquatic Sciences 74: 1362–1374. https://doi.org/10.1139/cjfas-2016-0306

Floerl O, Inglis GJ (2005) Starting the invasion pathway: the interaction between source populations and human transport vectors. Biological Invasions 7: 589–606. https://doi.org/10.1007/s10530-004-0952-8

Goldberg CS, Sepulveda A, Ray A, Baumgardt J, Waits LP (2013) Environmental DNA as a new method for early detection of New Zealand mudsnails (*Potamopyrgus antipodarum*). Freshwater Science 32: 792–800. https://doi.org/10.1899/13-046.1

Goldberg CS, Turner CR, Deiner K, Klymus KE, Thomsen PF, Murphy MA, Spear SF, McKee A, Oyler-McCance SJ, Cornman RS (2016) Critical considerations for the application of environmental DNA methods to detect aquatic species. Methods in Ecology and Evolution 7: 1299–1307. https://doi.org/10.1111/2041-210X.12595

Goldstien SJ, Dupont L, Viard F, Hallas PJ, Nishikawa T, Schiel DR, Gemmell NJ, Bishop JD (2011) Global phylogeography of the widely introduced North West Pacific ascidian *Styela clava*. PLoS One 6: e16755. https://doi.org/10.1371/journal.pone.0016755

Gollasch S (2002) The importance of ship hull fouling as a vector of species introductions into the North Sea.

Grey EK, Bernatchez L, Cassey P, Deiner K, Deveney M, Howland KL, Lacoursière-Roussel A, Leong SCY, Li Y, Olds B (2018) Effects of sampling effort on biodiversity patterns estimated from environmental DNA metabarcoding surveys. Scientific Reports 8: 8843. https://doi.org/10.1038/s41598-018-27048-2

Hadziavdic K, Lekang K, Lanzen A, Jonassen I, Thompson EM, Troedsson C (2014) Characterization of the 18S rRNA gene for designing universal eukaryote specific primers. PloS ONE 9: e87624. https://doi.org/10.1371/journal.pone.0087624

Hinlo R, Gleeson D, Lintermans M, Furlan E (2017) Methods to maximise recovery of environmental DNA from water samples. PloS ONE 12: e0179251. https://doi.org/10.1371/journal.pone.0179251

Holmes NJ (1968) Aspects of the biology of Styela clava Herdman. Southampton: University of Southampton.

Katsanevakis S, Wallentinus I, Zenetos A, Leppäkoski E, Çinar ME, Oztürk B, Grabowski M, Golani D, Cardoso AC (2014) Impacts of invasive alien marine species on ecosystem services and biodiversity: a pan-European review. Aquatic Invasions 9: 391–423. https://doi.org/10.3391/ai.2014.9.4.01

Kelly RP, O'Donnell JL, Lowell NC, Shelton AO, Samhouri JF, Hennessey SM, Feist BE, Williams GD (2016) Genetic signatures of ecological diversity along an urbanization gradient. PeerJ 4: e2444. https://doi.org/10.7717/peerj.2444

Lacoursière-Roussel A, Bock DG, Cristescu ME, Guichard F, Girard P, Legendre P, McKindsey CW (2012) Disentangling invasion processes in a dynamic shipping-boating network. Molecular Ecology 21: 4227–4241. https://doi.org/10.1111/j.1365-294X.2012.05702.x

Lacoursière-Roussel A, Howland K, Normandeau E, Grey EK, Archambault P, Deiner K, Lodge DM, Hernandez C, Leduc N, Bernatchez L (2018) eDNA metabarcoding as a new surveillance approach for coastal Arctic biodiversity. Ecology and Evolution. https://doi.org/10.1002/ece3.4213

Lacoursière-Roussel A, Rosabal M, Bernatchez L (2016) Estimating fish abundance and biomass from eDNA concentrations: variability among capture methods and environmental conditions. Molecular Ecology Resources 16: 1401–1414. https://doi.org/10.1111/1755-0998.12522

Leese F, Altermatt F, Bouchez A, Ekrem T, Hering D, Meissner K, Mergen P, Pawlowski J, Piggott JJ, Rimet F (2016) DNAqua-Net: Developing new genetic tools for bioassessment and monitoring of aquatic ecosystems in Europe.

Leray M, Knowlton N (2017) Random sampling causes the low reproducibility of rare eukaryotic OTUs in Illumina COI metabarcoding. PeerJ 5: e3006. https://doi.org/10.7717/peerj.3006

Li J, Lawson Handley LJ, Read DS, Hänfling B (2018) The effect of filtration method on the efficiency of environmental DNA capture and quantification via metabarcoding. Molecular Ecology Resources.

Longmire JL, Maltbie M, Baker RJ (1997) Use of" lysis buffer" in DNA isolation and its implication for museum collections. Occasional Papers of the Museum 163: 1–3. https://doi.org/10.5962/bhl.title.143318

Mächler E, Deiner K, Spahn F, Altermatt F (2015) Fishing in the water: effect of sampled water volume on environmental DNA-based detection of macroinvertebrates. Environmental Science & Technology 50: 305–312. https://doi.org/10.1021/acs.est.5b04188

McCutcheon C (2008) Port of Southampton. Amberley Publishing Limited, 160 pp.

Molnar JL, Gamboa RL, Revenga C, Spalding MD (2008) Assessing the global threat of invasive species to marine biodiversity. Frontiers in Ecology and the Environment 6: 485–492. https://doi.org/10.1890/070064

Munch K, Boomsma W, Huelsenbeck JP, Willerslev E, Nielsen R (2008) Statistical assignment of DNA sequences using Bayesian phylogenetics. Systematic Biology 57: 750–757. https://doi.org/10.1080/10635150802422316

Pearce F, Peeler E, Stebbing P (2012) Modelling the risk of the introduction and spread of non-indigenous species in the UK and Ireland. Report to Department for Environment, Food and Rural Affairs, Cefas Project code E5405W.

Pecl GT, Araújo MB, Bell JD, Blanchard J, Bonebrake TC, Chen I-C, Clark TD, Colwell RK, Danielsen F, Evengård B, Falconi L, Ferrier S, Frusher S, Garcia RA, Griffis RB, Hobday AJ, Janion-Scheepers C, Jarzyna MA, Jennings S, Lenoir J, Linnetved HI, Martin VY, McCormack PC, McDonald J, Mitchell NJ, Mustonen T, Pandolfi JM, Pettorelli N, Popova E, Robinson SA, Scheffers BR, Shaw JD, Sorte CJB, Strugnell JM, Sunday JM, Tuanmu M-N, Vergés A, Villanueva C, Wernberg T, Wapstra E, Williams SE (2017) Biodiversity redistribution under climate change: Impacts on ecosystems and human well-being. Science 355. https://doi.org/10.1126/science.aai9214

Piggott MP (2016) Evaluating the effects of laboratory protocols on eDNA detection probability for an endangered freshwater fish. Ecology and Evolution 6: 2739–2750. https://doi.org/10.1002/ece3.2083

R Core Team (2016) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna.

Renshaw MA, Olds BP, Jerde CL, McVeigh MM, Lodge DM (2015) The room temperature preservation of filtered environmental DNA samples and assimilation into a phenol-chloroform-isoamyl alcohol DNA extraction. Molecular Ecology Resources 15: 168–176. https://doi.org/10.1111/1755-0998.12281

Ruiz GM, Fofonoff PW, Carlton JT, Wonham MJ, Hines AH (2000) Invasion of coastal marine communities in North America: apparent patterns, processes, and biases. Annual Review of Ecology and Systematics 31: 481–531. https://doi.org/10.1146/annurev.ecolsys.31.1.481

Sassoubre LM, Yamahara KM, Gardner LD, Block BA, Boehm AB (2016) Quantification of environmental DNA (eDNA) shedding and decay rates for three marine fish. Environmental Science & Technology 50: 10456–10464. https://doi.org/10.1021/acs.est.6b03114

Seebens H, Gastner M, Blasius B, Courchamp F (2013) The risk of marine bioinvasion caused by global shipping. Ecology Letters 16: 782–790. https://doi.org/10.1111/ele.12111

Seymour M, Durance I, Cosby BJ, Ransom-Jones E, Deiner K, Ormerod SJ, Colbourne JK, Wilgar G, Carvalho GR, de Bruyn M (2018) Acidity promotes degradation of multi-species environmental DNA in lotic mesocosms. Communications Biology 1: 4. https://doi.org/10.1038/s42003-017-0005-3

Simpson TJ, Dias PJ, Snow M, Muñoz J, Berry T (2017) Real-time PCR detection of *Didemnum perlucidum* (Monniot, 1983) and *Didemnum vexillum* (Kott, 2002) in an applied routine marine biosecurity context. Molecular Ecology Resources 17: 443–453. https://doi.org/10.1111/1755-0998.12581

Spens J, Evans AR, Halfmaerten D, Knudsen SW, Sengupta ME, Mak SS, Sigsgaard EE, Hellström M (2017) Comparison of capture and storage methods for aqueous macrobial eDNA using an optimized extraction protocol: advantage of enclosed filter. Methods in Ecology and Evolution 8: 635–645. https://doi.org/10.1111/2041-210X.12683

Stat M, Huggett MJ, Bernasconi R, DiBattista JD, Berry TE, Newman SJ, Harvey ES, Bunce M (2017) Ecosystem biomonitoring with eDNA: metabarcoding across the tree of life in a tropical marine environment. Scientific Reports 7: 12240. https://doi.org/10.1038/s41598-017-12501-5

Stoeckle BC, Beggel S, Cerwenka AF, Motivans E, Kuehn R, Geist J (2017) A systematic approach to evaluate the influence of environmental conditions on eDNA detection success in aquatic ecosystems. PloS ONE 12: e0189119. https://doi.org/10.1371/journal.pone.0189119

Sylvester F, Kalaci O, Leung B, Lacoursière-Roussel A, Murray CC, Choi FM, Bravo MA, Therriault TW, MacIsaac HJ (2011) Hull fouling as an invasion vector: can simple models explain a complex problem? Journal of Applied Ecology 48: 415–423. https://doi.org/10.1111/j.1365-2664.2011.01957.x

Taberlet P, Bonin A, Zinger L, Coissac E (2018) Environmental DNA: For Biodiversity Research and Monitoring. Oxford University Press, Oxford.

Tsuji S, Ushio M, Sakurai S, Minamoto T, Yamanaka H (2017) Water temperature-dependent degradation of environmental DNA and its relation to bacterial abundance. PLoS ONE 12: e0176608. https://doi.org/10.1371/journal.pone.0176608

Turner CR, Barnes MA, Xu CC, Jones SE, Jerde CL, Lodge DM (2014) Particle size distribution and optimal capture of aqueous macrobial eDNA. Methods in Ecology and Evolution 5: 676–684. https://doi.org/10.1111/2041-210X.12206

Vasselon V, Domaizon I, Rimet F, Kahlert M, Bouchez A (2017) Application of high-throughput sequencing (HTS) metabarcoding to diatom biomonitoring: Do DNA extraction methods matter? Freshwater Science 36: 162–177. https://doi.org/10.1086/690649

Weltz K, Lyle JM, Ovenden J, Morgan JA, Moreno DA, Semmens JM (2017) Application of environmental DNA to detect an endangered marine skate species in the wild. PloS ONE 12: e0178124. https://doi.org/10.1371/journal.pone.0178124

Zar JH (1999) Biostatistical analysis. Prentice Hall, Upper Saddle River, New Jersey.

Supplementary material 1

Extraction protocols

Authors: Kristy Deiner, Jacqueline Lopez, Steve Bourne, Luke E. Holman, Mathew Seymour, Erin K. Grey, Anaïs Lacoursière-Roussel, Yiyuan Li, Mark A. Renshaw, Michael E. Pfrender, Marc Rius, Louis Bernatchez, David M. Lodge

Data type: methodology

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/mbmg.2.28963.suppl1

Supplementary material 2
Bioinformatic pipeline and thresholds
Authors: Kristy Deiner, Jacqueline Lopez, Steve Bourne, Luke E. Holman, Mathew Seymour, Erin K. Grey, Anaïs Lacoursière-Roussel, Yiyuan Li, Mark A. Renshaw, Michael E. Pfrender, Marc Rius, Louis Bernatchez, David M. Lodge
Data type: methodology
Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.
Link: https://doi.org/10.3897/mbmg.2.28963.suppl2

Supplementary material 3
Profiling tables for all libraries
Authors: Kristy Deiner, Jacqueline Lopez, Steve Bourne, Luke E. Holman, Mathew Seymour, Erin K. Grey, Anaïs Lacoursière-Roussel, Yiyuan Li, Mark A. Renshaw, Michael E. Pfrender, Marc Rius, Louis Bernatchez, David M. Lodge
Data type: methodology
Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.
Link: https://doi.org/10.3897/mbmg.2.28963.suppl3

Supplementary material 4
4_NTC
Authors: Kristy Deiner, Jacqueline Lopez, Steve Bourne, Luke E. Holman, Mathew Seymour, Erin K. Grey, Anaïs Lacoursière-Roussel, Yiyuan Li, Mark A. Renshaw, Michael E. Pfrender, Marc Rius, Louis Bernatchez, David M. Lodge
Data type: occurrences
Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.
Link: https://doi.org/10.3897/mbmg.2.28963.suppl4

Supplementary material 5
Alternative statistical model
Authors: Kristy Deiner, Jacqueline Lopez, Steve Bourne, Luke E. Holman, Mathew Seymour, Erin K. Grey, Anaïs Lacoursière-Roussel, Yiyuan Li, Mark A. Renshaw, Michael E. Pfrender, Marc Rius, Louis Bernatchez, David M. Lodge
Data type: statistical data
Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.
Link: https://doi.org/10.3897/mbmg.2.28963.suppl5